

TANULSÁGOK AZ ELKH-HRDA ADATREPOZITÓRIUM PILOT PROJEKTEK VÉGREHAJTÁSA ALAPJÁN

Holl András

MTA Könyvtár és Információs Központ

ORCID: [0000-0002-6873-3425](https://orcid.org/0000-0002-6873-3425)

Az Eötvös Loránd Kutatási Hálózat Titkársága a kutatási adatok kezelésére való felkészülés támogatására egyéves programot indított. A projekt szakmai támogatását a Research Data Alliance magyar tagozatának segítségével biztosította, a projekt technikai lebonyolítását és a költségvetés kezelését az MTA Könyvtár és Információs Központ végezte 2021-ben.

E program súlyponti részeként kutatásiadat-kezelési pilot projektek támogatására szolgáló pályázati kiírás jelent meg, melyre nyolc pályázatot adtak be az arra jogosult intézmények – az ELKH kutatóközpontjai, intézetei és kutatócsoportjai. Egy kivétellel minden pályázat támogatására mód nyílt, ugyan egy esetben csak csökkentett költségvetéssel.

A projektek egy részének eredményeiről készült beszámolók megtalálhatóak e kötetben.

Hasonló pilot projektek támogatására nem csupán itthon, de külföldön sem ismerünk példát. A nemzetközi – és immár hazai – kutatási pályázatokban mára már követelménnyé vált az adatkezelési tervek készítése, s így a kutatásiadat-kezelés költségei is elszámolhatóak. Kifejezetten kutatásiadat-kezelés megvalósítására azonban nem szoktak pályázatokat kiírni. A hazai kutatói társadalomnak csupán kis része – a legutolsó európai pályázati fordulóban támogatást nyert projektek, vagy a nemzetközi együttműködésben folytatott nagyprojektek résztvevői – találkozhatott a kutatásiadat-kezelés mára elfogadott FAIR

kritériumrendszerével. Az ELKH intézmények által megvalósított pilot projektek egyedülálló lehetőséget biztosítottak az adatkezelési gyakorlat fejlesztésére, a szabványok megismerésére, esetenként a szabványosítási folyamatba való bekapcsolódásra, a korábbi adatkezelési gyakorlatok megújítására.

Projektek

Támogatott projekt	Intézmény / szervezeti egység
Sokcsatornás, nagy téri felbontású in vivo elektrofiziológiai adatok archiválása*	Természettudományi Kutatóközpont, Kognitív Idegtudományi és Pszichológiai Intézet, Integratív Idegtudományi Kutatócsoport
A Társadalomtudományi Kutatóközpontban (illetve annak jogelődjében) végzett kutatások veszélyben lévő kutatási adatainak megóvása*	Társadalomtudományi Kutatóközpont, Kutatási Dokumentációs Központ
Régészeti rajzgyűjtemény kutatási adatainak feltárása és FAIR közreadása*	Bölcsészettudományi Kutatóközpont, Régészeti Intézet
Kis dózisoknál megfigyelhető hiperszenzitivitással és indukált sugárrezisztenciával kapcsolatos adatok gyűjtése és közzététele*	Energiatudományi Kutatóközpont, Energia- és Környezetbiztonsági Intézet, Környezetfizikai Laboratórium, Sugárbiofizikai Kutatócsoport
Fúziós kísérleti adatok tárolása és metaadatok tárolása a FAIR elveknek megfelelően	Energiatudományi Kutatóközpont, Fúziós Plazmafizika Laboratórium

* Ezen projektek beszámolóit megtalálhatók a kötetben.

Támogatott projekt	Intézmény / szervezeti egység
Az OpenBioMaps biológiai adatbázis keretrendszer publikus adatrepozitórium „láb” fejlesztésére	ELKH–DE Viselkedésökológiai Kutatócsoport
Funkcionális anyagok adatainak archiválása	Wigner Fizikai Kutatóközpont, Részecske- és Magfizikai Intézet, Nukleáris Anyagtudományi Osztály

A pilot projektek látványosan demonstrálták a kutatási adatok és kezelésük diverzitását. Nemhogy tudományterületek és -ágak között, de többnyire ugyanazon témában is alapvetően eltérő adattípusok fordulnak elő, melyek kezelése eltérő megközelítést kíván.

Az Energiatudományi Kutatóközpont Fúziós Plazmafizikai Laboratóriuma által megvalósított projektben a videodiagnosztikai és a nyalábemissziós spektroszkópiai mérésekben keletkező adatok mennyisége és feldolgozása is különböző. E projekt esetében a tárhelyszükséglet nagyságrendekkel haladta meg más projektekét – a tárolás és az adatmozgatás aspektus különbözteti meg a többitől. A nemzetközi szervezetek szabványosítási törekvéseibe való bekapcsolódás hangsúlyos eleme volt a munkának. Ugyancsak e projekt részeként valósult meg publikált cikkek kiegészítése mérési adatokkal.

A Régészeti Intézet projektje archív rajzdokumentáció digitalizálását, leírásának fejlesztését és adatrepozitóriumba helyezését célozta. Ennél a projektnél erőteljesen kidomborodott a történeti aspektus (nem a kutatott korszakokat, hanem a kutatás történetét tekintve): a hagyományos rajztár sok évtizedet felölelő, ugyanakkor folyamatos újrafelhasználási potenciállal bíró anyagainak digitális elérhetőségét és kereshetőségét kellett megalapozni, egyúttal lehetőséget adva a leíró adatok modern szempontok szerint történő gazdagítására, javítására. Ez a pályázat példázta a más hazai adatbázis (az Archeodatabase) szabványos, hierarchikus szövszedeteihez való alkalmazkodást.

Az Energiatudományi Kutatóközpont másik, dozimetriai projektje kis sugárdózisoknál megfigyelhető hiperszenzitivitás és indukált sugárrezisztencia modellezéséhez szükséges, a szakirodalomból gyűjtött adatok feldolgozását célozta. Ebben az esetben tehát rögtön megvalósult a korábbi, más kutatócsoportok által mért adatok újrafelhasználhatóvá tétele és újrafelhasználása: az összegyűjtött és közreadott adatok a modell javításán és ellenőrzésén túl további kutatások számára is hozzáférhetővé váltak.

Ismét másik oldalát mutatta meg a kutatásiadat-kezelésnek a TK KDK projektje. Ez esetben egy már régóta működő kutatási adatrepozitórium volt a pályázó, a megvalósított feladat pedig hanganyagok archiválása volt. Kiemelendő a hanganyagok kezelésének szoftveres megoldása, és a társadalomtudományok terén fontos adatvédelem, anonimizálás.

A Természettudományi Kutatóközpont projektje esetében is fontos tényező volt a nemzetközi adatleírási szabványokhoz való alkalmazkodás. Ennél a projektnél merült fel a publikációhoz társuló adatnyilvánossági követelmény is – a megvalósítás idején szembesültek a szerzők egy benyújtott közleményük bírálója kérésével, miszerint a felhasznált adatokat és az elemzésben alkalmazott kódot is tegyék elérhetővé.

A Wigner Fizikai Kutatóközpont kutatási programja keretében három (megjelent vagy elbírálás alatt lévő) közleményhez is elhelyeztek adatokat a Concorda-ban. A beszámolóban megjegyezték, hogy az adatrepozitóriumok a projektekben résztvevő, esetenként különböző intézményekből érkező kutatók közötti kommunikációban is fontos eszközök lehetnek: az adatok már a kísérletek során repozitóriumba kerülhetnek, és az arra jogosultaknak hozzáférhetőek lehetnek.

Újabb facettáját csillantotta meg a kutatásiadat-kezelésnek az MTA–DE Viselkedésokológiai Kutatócsoport projektje. Az OpenBioMaps egy kutatási célú adatbázis-infrastruktúra, melyhez adatrepozitálást elősegítő szoftveres megoldásokat fejlesztettek. Igen fontos a kutatáshoz használt eszközök (beleértve a szoftvereket és adatbázisokat) FAIR archiválást támogató funkciókkal való bővítése. Megfelelő

infrastruktúra nélkül a kutatók nem lesznek képesek a FAIR adatkezelés követelményei miatt megnövelt költség- és munkaigényeknek megfelelni.

Érdemes a program során előtérbe került sokféle követelményt felsorolni (még az ismétlés ódiomát is vállalva):

- nagy adatmennyiségek;
- kis kutatási projektek („Little Science”);
- archív anyagok;
- publikációkhoz kapcsolódó adatok;
- hazai szabványos nevezéktanok/szótárak használata;
- kereszthivatkozások hazai adatbázisokra;
- bekapcsolódás a nemzetközi szabványosítási folyamatba;
- média digitalizálási technológiák alkalmazása;
- adatbázisrendszerek kapcsolódásának kialakítása;
- korábbi adatok javítása;
- egyedi azonosítók használata.

Tanulságok

A HRDA tagjai/vezetősége köréből kikerült bíráló bizottság igen jó véleményt alakított ki a pilot eredményeiről. Lényeges eredmény volt, hogy olyan kutatókat és kutatócsoportokat is érzékenyíteni lehetett az adatarchiválás és a FAIR szempontrendszer követelményeivel, akik ezzel korábban nem találkoztak. Az ELKH épülő adatrepozitóriumára is lényeges volt a valós kutatói igényekkel való szembesülés, a korai kapcsolatépítés.

A projektbeszámolók alapján kiderül, hogy a pilot eredeti célkitűzésein túl is elért eredményeket:

„értékes információkhoz jutottunk mind az eredmények reprodukálhatóságát illetően, [...] mind pedig a tanulmány eredményeinek megbízhatóságát tekintve”¹

„egy másik [a pályázatban nem résztvevő] kutatócsoport [...] is megismerkedhetett a magyar adatrepozitóriummal [Concorda]”

„Hadd jegyezzük meg, hogy a kutatási adatok repozitóriumban való elhelyezése a kutatási projektek végrehajtása során, még az eredmények közzétevése előtt is egy nagyon hasznos eszköz lehet a kutatók kezében. Lehetőségeket nyújt, hogy az egyes adatsomagokhoz – privát URL-en keresztül – a közreműködő kollégák hozzáférjenek, ami nagymértékben megkönnyítheti a kutatók munkáját, főleg nagyobb adatmennyiségek esetén.”

Megállapíthatjuk, hogy

- i. a kutatásiadat-kezelés megfelelő méretű és biztonságú tárolóhelyek biztosítását igényli;
- ii. nemzetközi publikálás esetén egyre gyakrabban kötelező az archiválás és a hozzáférhetővé tétel;
- iii. a megfelelő kutatásiadat-kezelés munkaigényes;
- iv. az eredményes adatkezelés feltétele a megfelelő eszközök (szoftverek, protokollok, szolgáltatások) megteremtése;
- v. szaktudásra, támogatásra – adatgazdászok alkalmazására – van szükség.

Projektzáró, folytatás

A projektek 2021 decemberében lezárultak, a zárókonferencia 2022. január 18-án volt.²A pilot sikerét leginkább az tanúsítja, hogy az ELKH Titkársága folytatásként másfél éves futamidejű Adatrepozitórium Platform (ARP) projektet indított.

1 Az idézetek a projektbeszámolókból származnak.

2 <https://openaccess.mtak.hu/event/kutatasiadat-archivalasi-pilot-projektek-az-eotvos-lorand-kutatasi-halozathoz-tartozo-kutato kozpontokban-intezetekben-es-csoportokban/>